

**Kaplan C.S.<sup>1</sup>, Nassimova G.O.<sup>2</sup>, Brady H.E.<sup>3</sup>**

<sup>1</sup>Professor, doctor PhD, University of California, USA, Santa Barbara, e-mail: cskaplan@ucsb.edu

<sup>2</sup>Professor, doctor of Political Sciences, al-Farabi Kazakh National University,  
Kazakhstan, Almaty, e-mail: gulnar.nasimova@kaznu.kz

<sup>3</sup>Professor, doctor PhD, University of California, USA, Berkeley, e-mail: hbrady@berkeley.edu

**RESEARCH NOTE: METHODS IN STUDYING  
UNIVERSITY YOUTH IN KAZAKHSTAN**

The article presents new methods for the research of student youth in Kazakhstan and the basis of their implementation. The approaches used in the analysis of the political culture of student youth of Kazakhstan were developed in the framework of a research project supported by the Ministry of Education and Science of the Republic of Kazakhstan. The basic principles used in the construction of the survey sample and the development of the questionnaire, along with the determination of the number of students required in the sample needed to obtain the requisite information is described. Students at 10 institutions of higher education in Kazakhstan were surveyed with the number of respondents totaling 1161. The basic questions and problems encountered by the authors in the process of research are presented. The need for purposeful sampling is provided. The size of the sample makes possible the use of regression analysis. A description of how weights were created for the presentation of descriptive statistics is given. All of this made it possible to obtain data for determining and explaining the level of political activeness among students in Kazakhstan, their level of trust in political institutions, and the study of political values among student youth.

**Key words:** methods, sample, regression analysis, questionnaire, youth, students.

Каплан С.К.<sup>1</sup>, Насимова Г.О.<sup>2</sup>, Брэди Х.Е.<sup>3</sup>

<sup>1</sup>профессор, PhD докторы, Калифорния университеті,  
АҚШ, Санта Барбара қ., e-mail: cskaplan@ucsb.edu

<sup>2</sup>профессор, саяси ғылымдар докторы, әл-Фараби атындағы Қазақ ұлттық университеті,  
Қазақстан, Алматы қ., e-mail: gulnar.nasimova@kaznu.kz

<sup>3</sup>профессор, PhD докторы, Калифорния университеті,  
АҚШ, Беркли қ., e-mail: hbrady@berkeley.edu

**Зерттеу жазбасы: Қазақстандағы университеттік жастарды  
зерттеудің әдістері**

Берілген мақалада Қазақстандағы студент жастарды зерттеудің жаңа әдістері мен оны өткізудің негізгі кезеңдері ұсынылған. Бұл әдіс Қазақстан Республикасының Білім және ғылым министрлігінің ғылыми жобасы шеңберінде Қазақстандағы студент жастардың саяси мәдениетін талдауда қолданылды. Зерттеу жазбасында қажетті ақпаратты алу үшін сұралатын студенттер мөлшері, сауалнаманы сұрыптау мен әзірлемелерді құрудың негізгі қағидалары толық түсіндірілген. Қазақстандағы 10 жоғары оқу орындарының студенттері сұралып, респонденттер саны 1161 құрады. Зерттеу барысында авторлар ұшырастырған негізгі сауалдар мен мәселелер көрсетілді. Мақсатқа бағытталған сұрыптауды қолданудың қажеттілігі дәлелденді. Сұрыптау көлемі регрессивті талдау жүргізуге мүмкіндік берді. Мұның барлығы Қазақстандағы студенттердің саяси белсенділігін, олардың билік институттарына сенімі деңгейін, саяси құндылықтарын айқындау бойынша мәліметтер алуға мүмкіндік берді.

**Түйін сөздер:** әдістер, сұраптама, регрессиялық (regression) талдау, жастар сауалнама, студенттер.

Каплан С.К.<sup>1</sup>, Насимова Г.О.<sup>2</sup>, Брэди Х.Е.<sup>3</sup>

<sup>1</sup>профессор, доктор PhD, Университет Калифорнии, США, г. Санта-Барбара, e-mail: cskaplan@ucsb.edu

<sup>2</sup>профессор, доктор политических наук, Казахский национальный университет имени аль-Фараби, Казахстан, г. Алматы, e-mail: gulnar.nasimova@kaznu.kz

<sup>3</sup>профессор, доктор PhD, Университет Калифорнии, США, г. Беркли, e-mail: hbrady@berkeley.edu

### **Исследовательская записка: методы исследования университетской молодежи в Казахстане**

В данной статье представлены новые методы исследования студенческой молодежи в Казахстане и основные этапы их проведения. Данный подход применялся к анализу политической культуры студенческой молодежи Казахстана в рамках научного проекта Министерства образования и науки Республики Казахстан. В исследовательской записке дано подробное объяснение основных принципов построения выборки и разработки анкеты, определения количества опрашиваемых студентов для получения необходимой информации. Были опрошены студенты 10 высших учебных заведений Казахстана, число респондентов составило 1161. Показаны основные вопросы и проблемы, с которыми столкнулись авторы в процессе исследования. Обоснована необходимость применения целенаправленной выборки. Объем выборки дает возможность использовать регрессионный (regression) анализ. Все это позволило получить данные для определения политической активности студентов Казахстана, уровня их доверия к институтам власти, изучения политических ценностей студенческой молодежи.

**Ключевые слова:** методы, выборка, регрессионный анализ (regression), анкета, молодежь, студенты.

The research project «The Formation of Political Culture of Student Youth» led by Professor Gulnar Nassimova, Al-Farabi Kazakh National University, sought to explore many important themes connected with university students in Kazakhstan - the level of political knowledge of students and their interest in politics, the sources of news about politics, the forms of political participation, participation in student organizations and club activities, and level of trust among students in state institutions. These issues can be studied in a variety of ways using qualitative evidence, such as, interviews, focus groups and documents, and quantitative evidence, such as, aggregate data and surveys. With the participation of Professor Cynthia Kaplan, University of California, Santa Barbara who served as the senior foreign consultant on the project, it was decided that a survey of university students would be a valuable part of the research.

Surveys confront researchers with a series of serious practical considerations beyond the theoretical issues being explored. Who is the object of the study and how will they be selected in a manner which avoids bias? A large representative sample of the university student population might in theory be possible, but it would be very expensive and inefficient. As is well known in sampling theory, among the most important questions must be, what is the smallest group based on their characteristics about which you would like to draw statistically significant conclusions? Ultimately, this involves determining

the general sample size to be selected and the characteristics of the specific groups of interest. For example, this might include the regions of Kazakhstan, urban/rural residence, and ethnicity. We might have followed an approach which created clusters of universities which drew their students from the same areas and then randomly selected a subset for study. However, the knowledge of from where each university drew its students was unavailable. Furthermore, we would have to select a sufficient number of students with characteristics of interest from the homogeneous clusters. This granularity of information was unavailable.

Another issue is how we would select the actual students to be surveyed. Using a territorial approach, such as, a random walk to select place of residence followed by a quota for sex, ethnicity, and age would be inefficient. An alternative, which might limit the sample to students living in dormitories might create bias. Using lists of enrolled students from the universities selected within a cluster might be possible with the use of a random interval for selection of specific students, but such lists were not available. In short, with limited funds and incomplete data to create a sampling frame, the methods employed required creativity. What follows is how we addressed these issues in our research.

The Kazakhstan Youth Survey was created by Professors G.O. Nassimova and C.S. Kaplan and fielded in the fall/winter 2016 with a small supplement in the Spring 2017. The questionnaire was

written in Russian and translated into Kazakh. Professional back translation was used to ensure accuracy of meaning. The survey was self-administered lasting approximately 25 minutes. The research employs best practices in its approach to obtaining a university student population sufficiently large to permit statistical significance for causal analysis. Anonymity of students was stipulated and questionnaires were available in both the Kazakh and Russian languages. All data were entered at Al-Farabi Kazakh National University with those editing questionnaires switching questionnaires to do data entry in order to provide an extra check in order to avoid mistakes. Data were cleaned at the University of California, Santa Barbara. The number of survey respondents was 1,161.

*Sample Design: Purposeful Sample with an Oversample*

The population of interest to our study is university youth. Given financial constraints using a representative sample based on residence would be too expensive and inefficient. We chose an alternative approach to obtain a population of university youth—a purposeful sample. We recognized from the start that a purposeful sample presents some limits on inference, but our intent was to employ regression analysis as a means to explain variance in attitudes and behaviors, not descriptive statistics. With this in mind, two concerns dominated our approach to sampling. First was to obtain a sufficient number of students to allow for tests of statistical significance. Second was to be sufficiently inclusive to capture diversity among the student population. As in best practices, we did not select any student based on a factor which we would study (dependent variable) - attitudes or behavior. In addition, the distribution of the attitudes and behaviors of interest were for the most part unknown in the population which we wished to study.

Two types of characteristics were foremost in creating the sample - geographic and ethnic diversity. We think about these two factors as cells from a sampling frame aiming to have a sufficient number of respondents with these characteristics to allow for statistical analysis. Geographic diversity was addressed through choosing 10 universities located in different areas of Kazakhstan. After the survey data were entered, we analyzed where the students lived before they entered university. This allowed us to cluster regions by student origin, not simply analyze students based on the university which they attend. Thus, for each university we compared where the student was from and were able to make decisions about clusters based on the primary regions

of origin including both cities and rural areas for our subsequent analysis. This meant that we could test for the degree of homogeneity in clustered regions to see the degree to which attitudes and behaviors might differ had we included additional institutions from regions which were not among the universities in our sample. With this in mind, we used regional clusters of 7 areas based on those normally employed in representative samples and a simplified version having only 3 clusters. Ethnic diversity was of concern in creating a purposeful sample. Since attitudes and behaviors might differ related to ethnicity, we chose to create an oversample of ethnic Russians in order to have a sufficient number for statistical analysis. Students identifying with other ethnic groups were included in our initial sample, but we did not attempt to oversample them because there were too many national groups to do so and these groups did not constitute a single population about which inferences could be drawn.

### Selection of Students

At each university those conducting the survey were given a target number of students to have complete the survey. This target number was broken down by ethnic Kazakhs and Russians, although at no time was any student chosen based on ethnicity. Selection was based on the language in which a class was offered—Kazakh or Russian, not the ethnicity of students. This meant that students of all nationalities studying in Kazakh completed the survey and, of course, that Kazakhs who were studying in the Russian language completed surveys in classes in which Russian was the language of instruction. Only four Russians studied in classes in which Kazakh was the language of instruction, so almost all were in classes taught in Russian. Students of other nationalities studying in either Russian, or Kazakh and were included in our sample with their data entered. However, given the large number of ethnic groups and the small numbers of students in each group, we do not use these students in our analysis.<sup>1</sup>

We oversampled classes offered in Russian in order to create a quota sample aimed at increasing the number of ethnic Russians. We based the target number on two considerations. First, we sought a total number of ethnic Russian students which would allow for statistical analysis. Second,

<sup>1</sup> Given the diversity of these students, they do not constitute a single underlying population about which we can draw inferences.

since the ethnic Russian population might differ between the north of Kazakhstan and those in the south, particularly Almaty, we attempted to have equal numbers from the north and south. Again, no student was selected based on ethnicity. After questionnaires were completed those conducting the fieldwork were instructed to look at the answer to the question asking the ethnicity of the student. If the total number of ethnic Russians did not meet the target number needed for analysis, additional classes in the Russian language were selected to be included. Once the number of ethnic Russians specified in the target needed for statistical significance was met, the fieldwork stopped.

### Age

Consistent with the object of the study, our aim was to sample university students between the ages of 18 and 22. The average age of respondents was just a little over 19 years of age. This differs from many studies of youth in Kazakhstan which consider youth to include 14 to 29 year olds. Because we consider university age students a product of primary socialization which ends with the completion of secondary school, we exclude the youngest cohort who are unlikely to have developed their own interests and behaviors. It is well known that being away from home, or at university with their peers often leads to behavior and attitudes different from those when they were younger. Also, as is the practice in the United States, those younger than 18 years of age were excluded as stipulated by Human Subject Protocols.<sup>2</sup> We also did not want to include older individuals, 25-29 years of age, who might be influenced by jobs, marriage, and families which often exert a more conservative, or risk adverse set of attitudes. Our aim was to target those students at the start of their secondary socialization process, thought of as critical years.

### Universities

The 10 universities at which we fielded the survey were a convenience sample which included both private and public universities throughout Kazakhstan.<sup>3</sup> As part of assessing the external

<sup>2</sup> Those younger than 18 years of age require the consent of their parents.

<sup>3</sup> Al-Farabi Kazakh National University, Almaty, Women's Pedagogical University, Almaty, Bolashak, Kyzylorda, Korkit Ata, Kyzylorda, Ahmet Yassawi, Turkestan, NE Technical University, Oskemen, Eurasia National University, Astana, Karagandy State University, Karagandy, Alma University

validity and robustness of our findings, we compared our answers with those to equivalent questions which appeared on the World Value Survey in Kazakhstan 2011 (University of Michigan with fieldwork conducted by BISAM, Almaty). A sub-sample of respondents selected for age and level of education (N=400) was used from the World Value Survey in comparative analysis. On the equivalent items analyzed, we found our results were consistent with those of the World Value Survey.<sup>4</sup>

### Creating a Weighted Sample

Descriptive statistics from a non-representative sample, in general, should not be used for the purpose of external inference to the general population, although the data provide a perfectly acceptable basis for causal analysis using techniques such as Ordinary Least Squares regression, or Logit regression. Nevertheless, it is understandable that our colleagues were curious as to how the attitudes and behaviors found in our surveys are related to the general university student population. With this concern in mind, we consulted Professor Henry E. Brady at the University of California, Berkeley to provide a guide in creating a weighted sample that would resemble a representative sample. The resulting weighted data are used for descriptive statistics, but not in regression analysis. What follows is our approach to creating a weighted sample.<sup>5</sup>

### The Problem

We have a sample with given numbers of Kazakh and Russian students in each region. We want to reweight the numbers of Kazakhs and Russians to get a representative sample of Kazakh and Russian students for Kazakhstan. We know the number of Kazakhs and the number of Russians in each region.

---

(International Business Academy), KIMEP, Almaty. An attempt was made to include a predominantly Russian speaking university in the north of Kazakhstan, but it was unsuccessful.

<sup>4</sup> We also attempted to do this with data from the Ebert Fund's survey of Youth in Kazakhstan (2016). The results of the Ebert Fund's survey matching for age, education, and ethnicity departed from the World Value Survey and our own data. This caused us to examine the sample used. Due to anomalies found, we decided that this test was unreliable and would not contribute to a test of external validity.

<sup>5</sup> We present only the principles used, omitting the technical basis (equations) upon which our choices were based. Please contact Kaplan, if you wish to see the detailed logic employed in creating sample weights.

We know the number of students in each region, but not by their ethnicity.

*Step One – Adjusting for Ethnicity* For each region, we first adjust the sample numbers to get the right proportion of Kazakhs and Russians. We will assume, because there is no better assumption, that the ratio of Kazakh and Russian students in a region is the same as the ratio of Kazakh and Russian people.<sup>6</sup> To do this, we compute the weight for Kazakhs in each region (and for Russians).

Note that total observed sample size and total weighted sample size will almost surely be different. In fact, the total weighted sample size will be smaller than (or only in exceptional circumstances equal to) the total observed sample size since the sum of all people with weights that must be between zero and one will be less than the total observed sample size which essentially weights each observation with one. This does not cause any mathematical problems, but it should be noted that if just the weights are used the total weighted sample size will be smaller than the total observed sample size. It should also be understood that the weighted sample size could be adjusted to be bigger or smaller – the choice of the «correct» weighted sample size requires other considerations discussed below

*STEP TWO – Adjusting for number of students in each region*<sup>7</sup> -- Now, we want to adjust the weighted student sample sizes for each region with weights so that the students sampled are in the same ratios to one another as the number of students in the population are to one another.

Thus, we got values for the ratios of all the weights. We need one more piece of information to get the weights themselves. One approach is to simply set one weight, say the first one for the «first» region, to the number 1. That is the simplest procedure. This is what we did which is a liberal interpretation of how much information is in the sample

*STEP THREE – Dealing with Total Sample Size* -- Another approach to getting the weights other than setting the weight in one region to the value one is to require that the overall total weighted sample size  $N$  be equal to some number, perhaps the original total sample size. This places a constraint on the weights.

Note that for inferential purposes, the size of

the sample will have an impact (with larger samples implying more statistical significance for inferences) so that some thought should be given to what is used. Statisticians talk about the «effective» sample size for a weighted sample. It turns out that there is no one answer to what that should be.

*OTHER ISSUES – Regional versus Oblast Data* – This discussion has assumed that the oblast data have been aggregated into regions. We combined oblasts into regions based on creating homogeneous clusters of oblasts, similar to those created by survey firms in Kazakhstan. Then there are mechanical requirements of combining the number of Kazaks or Russians in the larger region from the oblast data or the number of students in the region from the oblast data. As has been noted, we focused on just Kazakhs and Russians.

Our survey research differs from earlier studies in its focus on university age students. Lacking data needed to create a classic sampling frame, we adopted an approach which would include a sufficiently large and diverse selection of students to allow for statistical analysis, in particular, the use of regression techniques. This meant that although we used a convenience sample of universities, we made sure that they were located in the different regions of Kazakhstan and we asked our respondents where they had lived before entering university. In addition, to geographic diversity, we realized that the random choice of students would not yield a sufficient number of ethnic Russian students for analysis. We therefore chose classes by language of instruction followed by a quota to increase the number of ethnic Russian students, since students studying in Russian might be Kazakh as well as from other nationalities. In setting our quotas for each university, we made sure that the number of self-identified Russian students was approximately equal from northern and southern (Almaty) Kazakhstan. Finally, in order to draw inferences about the general university student population in Kazakhstan from descriptive statistics, we created a weighted sample based on the distribution of university students by oblast in Kazakhstan and used oblast level data on ethnicity. Using these weights provided an approximation of a representative sample of university students in Kazakhstan.

We would like to thank the Ministry of Education and Science, Republic of Kazakhstan for its support of this research and Al-Farabi Kazakh National University for its assistance. We would also like to thank the University of California, Santa Barbara for helping to defer the costs of entering and cleaning survey data.

<sup>6</sup> This is an assumption we had to make because no data was available on the ethnic composition of students at specific institutions of higher learning.

<sup>7</sup> The Ministry of Education and Science makes public the number of students at institutions of higher education by oblast. We used these data for creating regional weights.